Do cosmic ray air showers initiate lightning?

Hare, B. M.; Dwyer, J. R.; Winner, L. H.; Uman, M. A.; Jordan, D. M.; Kotovsky, D. A.; Caicedo, J. A.; Wilkes, R. A.; Carvalho, F. L.; Pilkey, J. T.

Published in:
Journal of geophysical research-Atmospheres

DOI:
10.1002/2016JD025949

Publication date:
2017

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2016JD025949

Key Points:
- Cosmic ray air shower data correlated with lightning mapping array data provide no evidence for cosmic ray-initiated lightning
- Less than 5% of detected cosmic ray air showers could potentially have initiated a lightning flash
- We cannot exclude the possibility that lightning flashes are initiated by cosmic ray air showers

Correspondence to:
B. M. Hare,
bhare8972@hotmail.com

Citation:
Hare, B., et al. (2017), Do cosmic ray air showers initiate lightning?: A statistical analysis of cosmic ray air showers and lightning mapping array data, *J. Geophys. Res. Atmos.*, *122*, 8173–8186, doi:10.1002/2016JD025949.

Received 16 SEP 2016
Accepted 6 JUL 2017
Accepted article online 21 JUL 2017
Published online 10 AUG 2017

Do cosmic ray air showers initiate lightning?: A statistical analysis of cosmic ray air showers and lightning mapping array data

B. M. Hare[1], J. R. Dwyer[2], L. H. Winner[3], M. A. Uman[4], D. M. Jordan[4], D. A. Kotovsky[4], J. A. Caicedo[4], R. A. Wilkes[4], F. L. Carvalho[4], J. T. Pilkey[4], T. K. Ngin[4], W. R. Gamerota[4], and H. K. Rassoul[5]

[1]KVI-Center for Advanced Radiation Technology, University of Groningen, Groningen, Netherlands, [2]Department of Physics and Space Science Center (EOS), University of New Hampshire, Durham, New Hampshire, USA, [3]Department of Statistics, University of Florida, Gainesville, Florida, USA, [4]Department of Electrical Engineering, University of Florida, Gainesville, Florida, USA, [5]Department of Physics, Florida Institute of Technology, Melbourne, Florida, USA

Abstract It has been argued in the technical literature, and widely reported in the popular press, that cosmic ray air showers (CRASs) can initiate lightning via a mechanism known as relativistic runaway electron avalanche (RREA), where large numbers of high-energy and low-energy electrons can, somehow, cause the local atmosphere in a thundercloud to transition to a conducting state. In response to this claim, other researchers have published simulations showing that the electron density produced by RREA is far too small to be able to affect the conductivity in the cloud sufficiently to initiate lightning. In this paper, we compare 74 days of cosmic ray air shower data collected in north central Florida during 2013–2015, the recorded CRASs having primary energies on the order of 10^{16} eV to 10^{18} eV and zenith angles less than 38°, with Lightning Mapping Array (LMA) data, and we show that there is no evidence that the detected cosmic ray air showers initiated lightning. Furthermore, we show that the average probability of any of our detected cosmic ray air showers to initiate a lightning flash can be no more than 5%. If all lightning flashes were initiated by cosmic ray air showers, then about 1.6% of detected CRASs would initiate lightning; therefore, we do not have enough data to exclude the possibility that lightning flashes could be initiated by cosmic ray air showers.

1. Introduction

Gurevich et al. [1992] first described the phenomenon now known as Relativistic Runaway Electron Avalanche (RREA). RREA starts with a runaway electron that has enough initial energy that it continues to gain energy while traversing an electric field (such as is present in a thunderstorm) despite energy losses primarily due to ionization. *Gurevich et al.* [1992] showed that when such a runaway electron undergoes Møller scattering, it can excite new free electrons, some of which will then have enough energy to also run away. This shower of high-energy electrons exponentially increases in number as it traverses the electric field. The initial seed electrons for RREAs can be supplied by several different sources, including natural radioactivity and cosmic ray air showers (CRAS). *Gurevich et al.* [1992, 1999, 2004] and *Gurevich and Zybin* [2005] postulated that RREA, seeded by CRASs, can initiate a lightning flash. *Marshall et al.* [1995] used the measurements from multiple balloon-borne electric field soundings of thunderstorms to support the view that CRASs, via RREA, initiate lightning since they were able to show that the electric field in a thunderstorm rarely increases over the critical value necessary for RREA. *Solomon et al.* [2001] compared this CRAS initiation hypothesis to the primary alternative hypothesis that lightning flashes are initiated by the enhanced local electric field provided by hydrometers, and they concluded that it is more likely that CRASs are initiating lightning than are hydrometeors. *Gurevich et al.* [2009] claim to have directly detected an intracloud flash that was initiated by a cosmic ray air shower. *Chonis* [2009] compared data from the National Lightning Detection Network to cosmic ray data archived by the National Geophysical Data Center and found a correlation between monthly lightning flash rates and monthly galactic cosmic ray flux during the winter season. Finally, *Gurevich and Karashtin* [2013] combined both the hydrometeor initiation hypothesis and the cosmic ray air shower initiation hypothesis to

argue that hydrometeors in thunderstorms could enhance the electric field created by RREA due to cosmic ray air showers enough to allow cosmic ray air showers with primary energies between 10^{11} eV to 10^{12} eV to initiate lightning flashes.

Dwyer [2005, 2007, 2010], *Babich et al.* [2011], and *Babich et al.* [2012] have shown that while RREA seeded by CRASs can, over time, create regions of enhanced electric field, it cannot directly initiate lightning because the low-energy electrons produced by RREA are far too diffuse to significantly affect the conductivity of a thunderstorm. *Milikh and Roussel-Dupré* [2010] argued that RREA produced an anomalous growth of low-energy electrons, but this was shown by *Dwyer and Babich* [2011] not to be the case. *Rison et al.* [2016] examined the initial breakdown of a number of lightning flashes with a lightning mapping array and a lightning interferometer and found that all the lightning flashes were initiated by fast positive breakdown, and so could not have been initiated by RREA.

In the present paper we use a statistical model to compare the time of CRASs detected by a 0.25 km^2 array with the time of lightning mapping array data sources occurring within 8 km of the International Center for Lightning Research and Testing (ICLRT). In section 2 we introduce the ICLRT cosmic ray and LMA networks. In section 3.1 we empirically find the distribution of LMA sources during times without CRASs and use Fisher's Method to show that there is no statistical difference between this empirical distribution and the distribution of LMA sources immediately following CRASs on time scales of 2 ms, 5 ms, 10 ms, and 25 ms. In section 3.2 we show that if the detected CRASs were initiating single LMA sources more than 5% of the time, then we would have found a statistically significant result. Since we do not have any statistical significance, we conclude that less than 5% of detected CRASs could have initiated LMA sources. We discuss these results further in section 4 and conclude in section 5.

2. Instrumentation

2.1. Cosmic Ray Air Shower Measurements

The International Center for Lightning Research and Testing (ICLRT) is a lightning research facility in north central Florida, 7.5 km east of Starke, Florida, where both natural and artificially initiated (triggered from natural thunderstorms using the rocket-and-wire technique) lightning flashes are studied. The ICLRT operates a network of eight plastic particle scintillators that are used to detect the muons associated with cosmic ray air showers. The cosmic ray network is described thoroughly by *Hill* [2012]. A brief summary follows. Each plastic scintillator has an area of 1 m^2. The eight scintillators cover an area of about a quarter of a square kilometer. Each plastic scintillator is monitored by two photomultiplier tubes, whose outputs are summed together and transmitted to a central location via fiber optic cabling. At the central location, the data from the eight scintillators are stored on digital oscilloscopes that are triggered whenever four or more of the scintillators detect an energetic particle or particles (mostly muons) from an air shower within a 5 μ s window. The network was activated whenever storms were near the ICLRT.

In order to find the range of cosmic ray primary energies to which our cosmic ray network is sensitive, we integrated an empirical equation for the intensity of cosmic ray flux weighted by the probability to detect the cosmic ray air showers over energy, solid angle, and area, as given in equation (1):

$$\text{detection rate} = \int P(E,\theta,\phi,\rho,\Phi)\, I(E)\, \mathrm{d}E\, \mathrm{d}\Omega\, \mathrm{d}r^2 \tag{1}$$

where $P(E,\theta,\phi,\rho,\Phi)$ is the probability for the ICLRT cosmic ray network to detect a CRAS of primary energy E, zenith angle of the cosmic ray primary θ, azimuth angle of the cosmic ray primary ϕ, horizontal distance between the ICLRT network and first interaction of the cosmic ray primary ρ, and azimuth angle between the first interaction and the ICLRT network Φ. The integral in equation (1) is taken over the energy of the primary particle dE, solid angle observed by the detector dΩ, and area of the detector dr^2. $I(E)$ is the intensity of the cosmic ray flux versus primary energy, shown in equation (2) [*Gaisser*, 1990].

$$I(E) = 1.8\times 10^4 \left(\frac{E}{1\ \text{GeV}}\right)^{-2.7} \frac{1}{S\ \text{GeV m}^2\ \text{Sr}} \tag{2}$$

We approximate the probability of detecting a cosmic ray air shower using a very simple model. First, we approximate the total number of particles on the ground with equation (3) [*Gaisser*, 1990],

$$N_g = S_0(E)\frac{E}{\epsilon}\frac{p}{p+1}e^p \left(\frac{X}{X_{\max}-\lambda}\right)^{p+1} \exp\left(-\frac{X}{\lambda}\right) \tag{3}$$

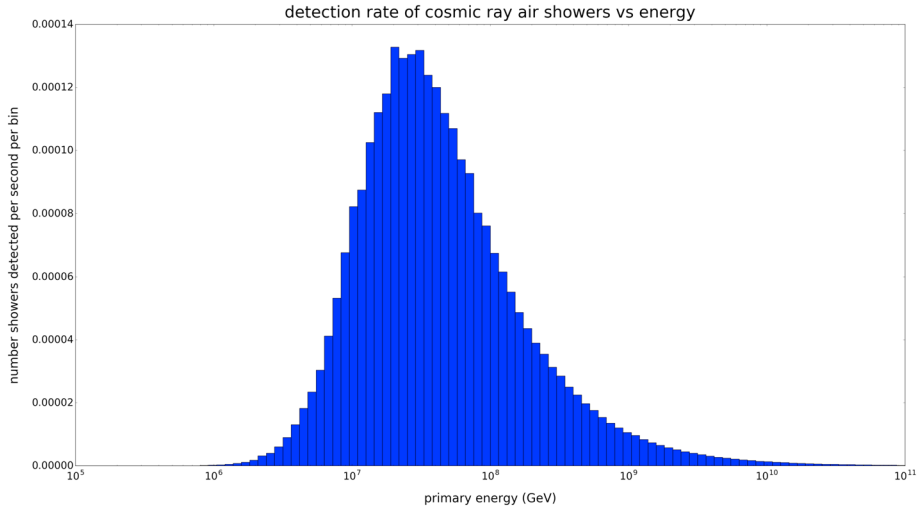

Figure 1. The approximate rate of cosmic ray air shower detections versus energy.

where $p = \frac{X_{\max}}{\lambda} - 1$, X is the slant depth at ground, which is dependent on the zenith angle of the CRAS primary, $X_{\max}$ is the slant depth of the shower maximum, which is given by $X_0' \ln(\frac{E}{\epsilon})$, λ is 70 g/cm^2, ϵ is 0.074 GeV. X_0' is 34.5 g/cm^2, and $S_0(E)$ is $0.04 \times (1.0 + 0.0217 \times \ln(\frac{E}{10^6 \text{ GeV}}))$.

We can then relate the density of particles on the ground to the total number of particles using a modified Nishimura-Kamata-Greisen (NKG) formula, shown in equation (4) [*Gaisser*, 1990]:

$$\rho(r) = \frac{N_g}{2\pi r_1^2 C(s)} (r/r_1)^{(s-2)} (1 + r/r_1)^{(s-4.5)} (1 + 0.088 \times r/r_1) \tag{4}$$

where $C(s)$ is given by

$$C(s) = B(s, 4.5 - 2s) + 0.088B(s + 1, 3.5 - 2s) \tag{5}$$

$B(x, y)$ is the beta function, $s = 1.25$, r_1 is the Mollier radius, which is 78 m at sea level, r is the distance between the detector and the cosmic ray air shower core, and $\rho(r)$ is the density of particles on the ground. Note that ρ in equation (4) is different from the ρ in equation (1).

Using equation (4), we found the density and expected number of particles incident on each of the ICLRT scintillators. The Poisson distribution was used to obtain the probability that there was one or more particles incident on each detector. Next, we found the probability that four or more scintillators detected a particle, assuming that if a particle is incident on a detector, then the scintillator will detect the particle. Finally, equation (1) was integrated using a Monte Carlo technique.

Figure 1 shows the results binned across energy, demonstrating that the detected CRASs have primary energies between about 10^{16} eV and 10^{18} eV. The ICLRT cosmic ray network is sensitive to cosmic rays with energies above 10^{18} eV, but these cosmic rays are too few to affect our statistical analysis. Figure 2 shows the results binned across the zenith angle of the cosmic ray air showers, demonstrating that 90% of the CRASs detected by the ICLRT have zenith angles less than 38°. Obviously, the method we use here is highly approximate. We can compare the results to LORA, which is an air shower array installed at the center of LOFAR, a large radio telescope in the Netherlands [*Thoudam S. et al.*, 2014, 2015]. LORA is similar to the ICLRT cosmic ray network. It uses 20 plastic scintillators, slightly over twice as many as the ICLRT, over an area with a diameter of about 320 m, which is a smaller physical area than the ICLRT. LORA is sensitive to cosmic rays with primary energies above 10^{16} eV [*Thoudam S. et al.*, 2014, 2015], indicating that the results of our simulation are reasonable. Furthermore, our simulation gives a total CRAS detection rate of 0.0027 detections/s, which is close to the actual rate of CRAS detections at the ICLRT (0.0060 detections/s).

2.2. Lightning Mapping Array

The ICLRT has an eight-station Lightning Mapping Array (LMA) that can locate in 3-D the sources of electrical breakdown in the thundercloud. Typically, these sources are at front of propagating lightning leader channels. The LMA is described in detail in *Pilkey* [2014]. Each LMA station consists of a VHF (around 70 MHz) antenna,

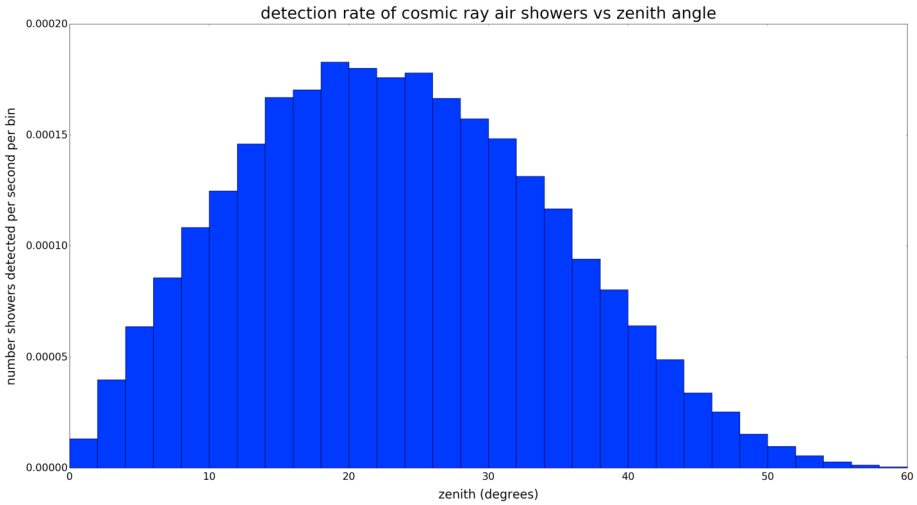

Figure 2. The approximate rate of cosmic ray air shower detections versus zenith angle.

computer, cell antenna, and modem for Internet access, batteries, and a GPS antenna. Signals from the VHF antenna passes through a preamplifier, a VHF channel 4 (66–72 MHZ) band-pass filter, and a log amplifier. The computer analyzes this signal in 10 μ s intervals looking for the time of the peak signal during that interval. If the peak signal amplitude is greater than a certain threshold, then the computer will record the power and time of the signal peak. The threshold floats up and down over time and is set by the computer so that the station records about 1000 signal peaks a second. These data are then saved to on-site drives and regularly collected from each station by hand. A TOA algorithm, using a point source model, is used to find LMA sources in 3-D and time. The timing uncertainty of the ICLRT LMA has been estimated to be 30–40 ns [*Pilkey et al.*, 2014]. The goodness of fit for each LMA source is calculated via a reduced chi-square value. The LMA stations are arranged around the ICLRT with baselines on the order of kilometers.

Since the ICLRT is only sensitive to CRASs with zenith angles less than 38 degrees, and *Pilkey et al.* [2014] showed that thunderstorm charge regions around the ICLRT have altitudes around 10 km and lower, then the CRASs detected by the ICLRT pass through thunderstorm charge regions, when one is present, within a 8 km horizontal distance of the ICLRT. For this reason, the LMA data used in this paper have been filtered to include only LMA sources that are within 8 km of the cosmic ray network. The LMA data have also been filtered to only include sources that have reduced chi-square values less than 1 and have a minimum of 6 stations contributing to each LMA source. This level of filtering is stringent enough that there are very few noise sources during periods of time where there is no electrical activity in the atmosphere. The LMA data was processed in 2 second windows around the time of each CRAS.

3. Statistical Analysis

3.1. Initial Analysis

In order to investigate a possible connection between CRASs and lightning initiation, we will statistically compare our detected CRASs to LMA sources. This, however, is difficult because most LMA sources during a lightning flash will be due to leader propagation and not lightning initiation. Furthermore, it is very possible that the LMA system will not detect the initial breakdown at all and will only detect the subsequent leader.

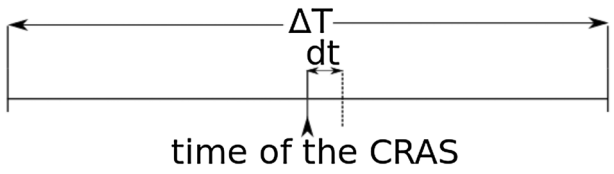

Figure 3. A graphical representation of when a LMA source is considered temporally coincident with a CRAS or in the background of a CRAS. If a LMA source is d*t* after the CRAS, then it is temporally coincident with that CRAS, or if the LMA source is within $\pm\Delta T/2$ of the time of the CRAS, then it is considered in the background of the CRAS. The background sources exclude the coincident sources.

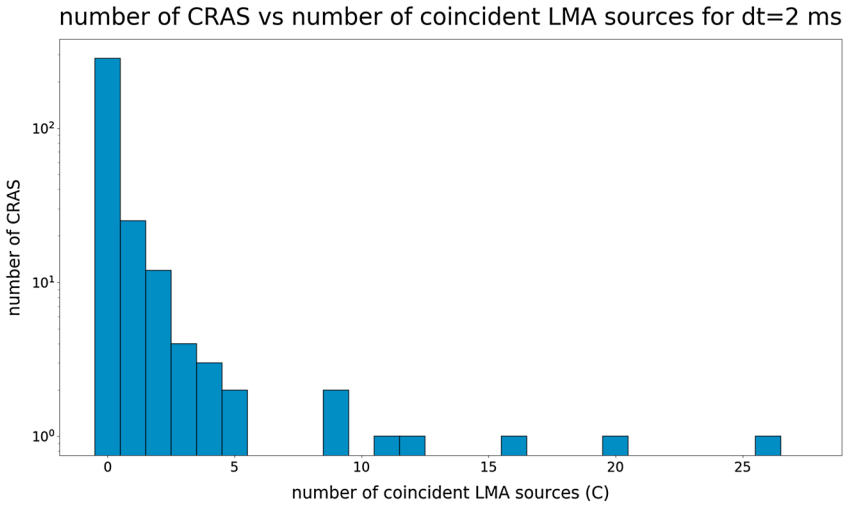

Figure 4. Histogram of the number of CRASs versus number of LMA sources coincident with the CRAS for dt = 2 ms.

For this reason, we will consider the scenario that a CRAS propagates through a thunderstorm region and initiates a lightning flash via some physical mechanism. The precise mechanism is not important in our study. This initial breakdown may or may not be detected by the LMA, but it will evolve into a leader which will be detected by the ICLRT LMA. In this way, a CRAS may initiate a lightning flash but we will say that a CRAS "produces" LMA sources, since those LMA sources may only be indirectly due to the CRAS. Thus, we will investigate if CRASs can produce LMA sources. To this end, we will compare the distribution of LMA sources immediately after CRAS to a distribution of LMA sources during times where there were no detected CRAS. Since we do not know precise time scale of how long it should take a CRAS to produce a LMA source, we will investigate multiple time scales in between the time scale of initial breakdown and the time scale of leader propagation, which is less than 2 ms and about 35 ms, respectively [*Rakov and Uman*, 2003].

In order to compare CRASs detected at the ICLRT to the LMA data, we first observed the number of coincident LMA sources that occurred in a period of time of length dt after each CRAS, which will be referred to as C_i for the ith cosmic ray air shower. We also observed the number of background LMA sources that occur in some time $\pm\Delta T/2$ around the CRAS, which will be referred to as B_i for the ith CRAS. The B_i background counts exclude the C_i coincident counts. The dt describes the time length of the process in which we are interested, and ΔT

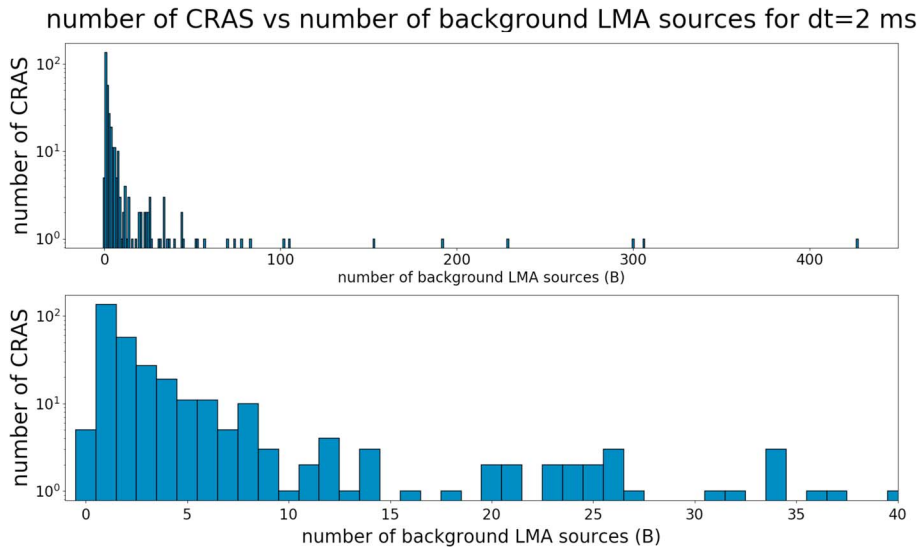

Figure 5. (top) Histogram of the number of CRASs versus number of LMA sources in the background of the CRAS for dt = 2 ms. (bottom) Same data as in Figure 5 (top) but is zoomed-in to show more detail for low number of background counts.

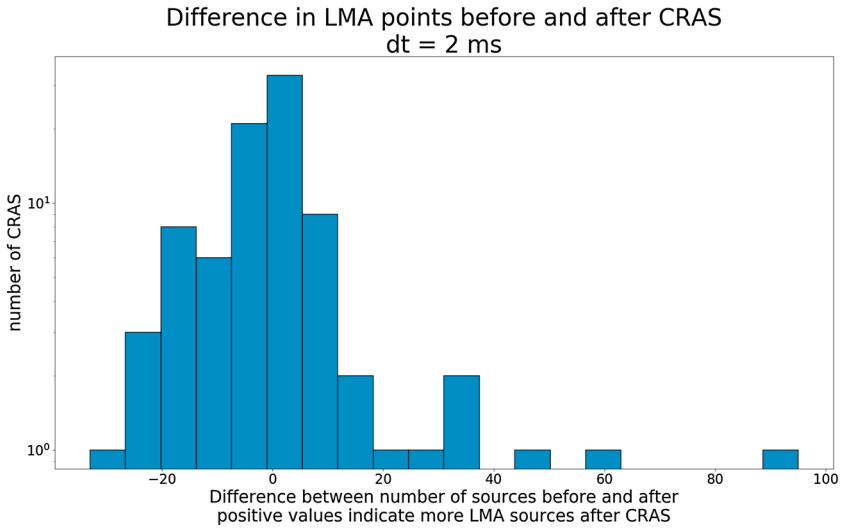

Figure 6. Histogram of number of CRAS versus differences between LMA sources before and after each CRAS for dt = 2 ms. Positive differences indicate more LMA sources after the CRAS.

is required to be much larger than dt. Here ΔT will always be chosen as 20 times dt. Figure 3 illustrates the relationship between dt, ΔT, and the time associated with the CRAS. The reason we record both the number of coincident LMA sources, C_i, and the number of background LMA sources, B_i, is so that we can find if C_i follows a different distribution between times immediately after a CRAS and times when there are no CRASs. We record B_i so that we can compare the distributions of C_i between similar lightning processes. That is, if a CRAS occurs during a time when there is no lightning, then B_i may be 0 for that CRAS. Then, a C_i of 1 or 2 could potentially be statistically significant. But if the CRAS occurs during the middle of a lightning flash, then B_i may potentially be large and a C_i of 1 or 2 may be exactly what is expected. Therefore, we will only compare the distributions of C_i to times that have the same B_i. Lastly, since we are only interested in CRASs during thunderstorms, we exclude CRASs that do not have at least one background or coincident LMA source. Figure 4 shows a histogram of number of CRASs versus number of coincident LMA counts for dt = 2 ms. Figure 5 shows a histogram of number of CRAS versus background LMA counts for dt = 2 ms. Figures 6–9 show a histogram of number of CRAS binned by difference between LMA sources before and after each CRAS. Positive differences mean that there are more LMA sources after the CRAS than before.

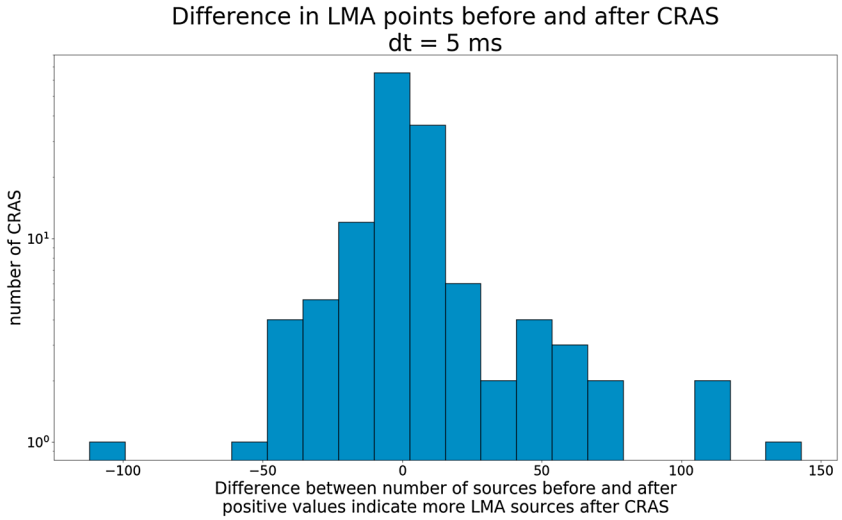

Figure 7. Histogram of number of CRAS versus differences between LMA sources before and after each CRAS for dt = 5 ms. Positive differences indicate more LMA sources after the CRAS.

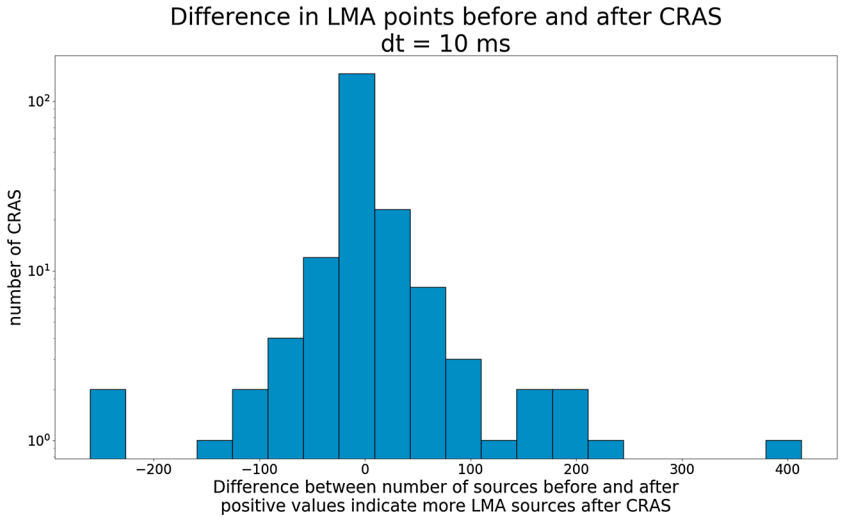

Figure 8. Histogram of number of CRAS versus differences between LMA sources before and after each CRAS for dt = 10 ms. Positive differences indicate more LMA sources after the CRAS.

After measuring C_i and B_i for each CRAS, we then found the distribution that C_i, given B_i, should follow if CRASs do not produce LMA sources. This was done by sampling a large number of random times and binning LMA sources into dt and ΔT bins. The times were chosen uniformly across all processed LMA data. We excluded any samples that contained a CRAS in the ΔT bin. We will refer to the resulting empirical probability mass function of the number of counts during dt (C), given the number of counts during ΔT (B), as $f(C|B)$. As before, the B counts in the ΔT bin exclude the C counts in the dt bin. Note that the LMA counts are not independent, and so $f(C|B)$ cannot be modeled by the binomial distribution.

Using our empirical distribution function, we can find a p value for each of our CRAS observations: $p_i = P(C \geq C_i|B = B_i) = \sum_{C=C_i}^{C=\infty} f(C|B_i)$. Figure 10 shows $-2\ln(p)$ as a function of C for $B = 1$ and $B = 6$, for dt = 2 ms. The challenge is combining the p values into a single statistic. One common method for combining p values is Fisher's Method [*Fisher*, 1973], given in equation (6)

$$\mathrm{TS} = \sum_{i=1}^{N} -2\ln(p_i) \tag{6}$$

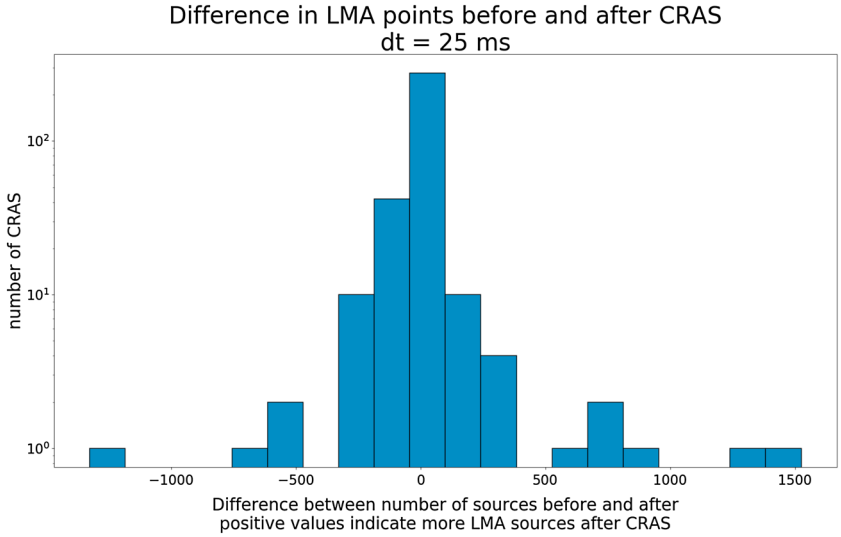

Figure 9. Histogram of number of CRAS versus differences between LMA sources before and after each CRAS for dt = 52 ms. Positive differences indicate more LMA sources after the CRAS.

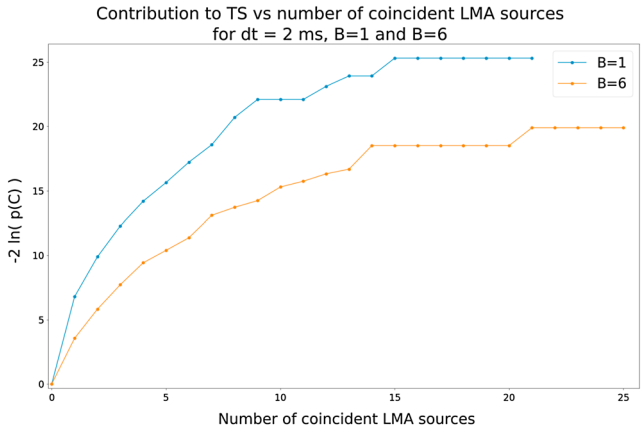

Figure 10. Contribution to TS, $-2\ln(p(C))$, as a function of number of coincident LMA sources when $B = 1$ and $B = 6$. For d$t = 2$ ms.

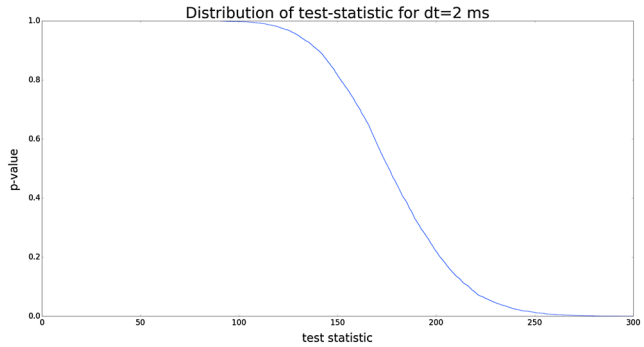

Figure 11. Distribution of the test statistic, TS, for d$t = 2$ ms. The y axis is the p value, $P(\text{ts} \geq \text{TS})$, and x axis is TS.

Table 1. The Number of Analyzed CRAS, *P* value, and TS at Each Analyzed Time Scale

	Number CRASs	*P* Value	TS
d$t = 2$ ms	325	0.76	155.1
d$t = 5$ ms	501	0.96	247.0
d$t = 10$ ms	651	0.76	429.6
d$t = 25$ ms	936	0.25	783.5

Table 2. The Average, Standard Deviation, Maximum, and Minimum of the Contribution to TS Made By Individual CRAS

	Average	Standard Deviation	Maximum	Minimum
d$t = 2$ ms	0.48	1.48	10.81	0.00
d$t = 5$ ms	0.49	1.61	14.09	0.00
d$t = 10$ ms	0.66	1.76	12.08	0.00
d$t = 25$ ms	0.84	1.93	14.34	0.00

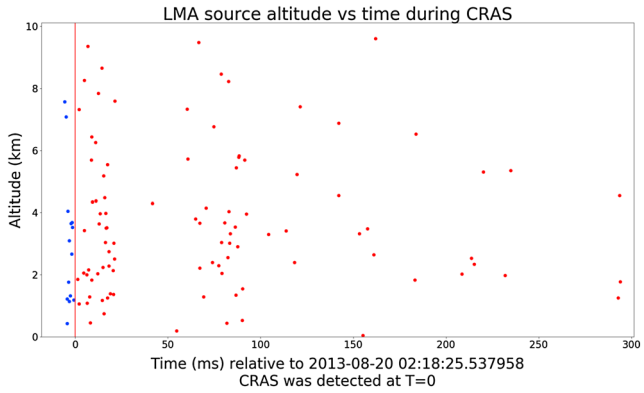

Figure 12. LMA sources around the time of the CRAS with the largest contribution to TS over all time scales. Each point is an LMA source with altitude along the *y* axis and time relative to the CRAS along the *x* axis. The blue LMA sources came before the CRAS and the red LMA sources came after the CRAS. The red bar shows the time the CRAS was detected, at $T = 0$.

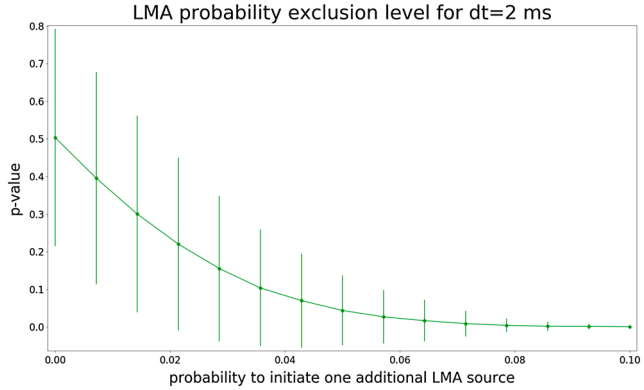

Figure 13. Monte Carlo simulation of our statistical model for d$t = 2$. The points show the mean *p* value, and the bars show the standard deviation of the *p* value. Enough Monte Carlo runs were performed that the standard error of the mean is smaller than 10^{-5}.

Table 3. The Average, Standard Deviation, Maximum, and Minimum of TS for 1000 Runs With Randomized CRAS Times

	Average	Standard Deviation	Maximum	Minimum
d$t = 2$ ms	173	31	270	86
d$t = 5$ ms	305	40	478	195
d$t = 10$ ms	454	48	634	288
d$t = 25$ ms	746	58	932	570

Table 4. The Average, Standard Deviation, Maximum, and Minimum of the *P* Values for 1000 Runs With Randomized CRAS Times

	Average	Standard Deviation	Maximum	Minimum
d$t = 2$ ms	0.50	0.29	1.00	0.00
d$t = 5$ ms	0.51	0.29	1.00	0.00
d$t = 10$ ms	0.49	0.29	1.00	0.00
d$t = 25$ ms	0.45	0.29	1.00	0.00

Table 5. The Average, Standard Deviation, Maximum, and Minimum of the Number of Participating CRASs for 1000 Runs With Randomized CRAS Times

	Average	Standard Deviation	Maximum	Minimum
dt = 2 ms	311.6	14.1	354	272
dt = 5 ms	479.9	16.3	541	427
dt = 10 ms	642.8	17.4	699	598
dt = 25 ms	934.4	17.0	990	878

where, if p_i is a continuous, uniformly distributed p value, then TS, the test statistic, would be a chi-square random variable with $2N$ degrees of freedom. This, however, is not the case for our data since $f(C|B)$ is a discrete distribution, and therefore our p values will not follow a uniform distribution and so the natural log of our p values will not be chi-square distributed.

In order to apply equation (6) we used a Monte Carlo simulation to find the distribution that TS actually follows. In each run of this Monte Carlo simulation, we looped over every cosmic ray air shower and sampled C from $f(C|B_i)$, found the associated p value, and then calculated TS using equation (6). Using this method we sampled the distribution of TS 6000 times for each dt. Figure 11 shows the results of this simulation for dt = 2 ms. The distribution of TS is similar for different values of dt. Lastly, we can use our distribution of TS to calculate a final p value for actual CRAS data. Under the null hypothesis, that CRAS do not produce LMA sources, the final p value will be a random uniform number between 0 to 1. However, if there are more LMA sources immediately after CRASs than can be explained by $f(C|B)$, then the final p value will become small. If the final p value is small enough, smaller than an alpha value that we set at 0.05, then we can exclude the null hypothesis that CRASs do not initiate LMA sources. The value 0.05 is a normally adopted significance level.

For time spans of dt = 2 ms, 5 ms, 10 ms, and 25 ms, we found final p values of 0.767, 0.962, 0.764, and 0.247, respectively. At these time scales, there were 325, 501, 651, and 936 cosmic ray air showers with at least one associated LMA source. These values, along with TS, are reported in Table 1. There were a total of 4884 detected CRASs. All of these p values are larger than our alpha value of 0.05, and therefore, we cannot exclude our null hypothesis that CRASs do not produce LMA sources.

In order to demonstrate the typical contribution that individual CRASs make to the total test statistic, TS, we also report the average, standard deviation, maximum, and minimum of $-2\ln(p_i)$ for each dt time scale in Table 2. Notice that out of all four time scales, the largest contribution to TS was 14.34. Figure 12 shows LMA sources of this event, with the LMA sources plotted by altitude (y axis) and time (x axis). The time is in units of ms from the time of the CRAS in question. The LMA sources colored blue happened before the CRAS, and the LMA sources colored red happened after the CRAS. The red bar shows the time of the CRAS, at $T = 0$. This figure is discussed more in section 4.

3.2. Exclusion Level Analysis

Our simple statistical analysis does not provide any evidence that CRASs produce LMA sources. However, it also does not provide any evidence that CRASs never produce LMA sources. In order to set an exclusion level on the maximum percent that CRASs could be producing LMA sources, we need to make an assumption on the distribution of LMA sources ostensibly made by cosmic ray air showers. The simplest assumption we can make is that all CRASs during a thunderstorm have an equal probability of producing precisely one extra LMA source. Obviously this assumption is not entirely correct, some CRASs could have a higher probability and some have a lower probability of producing different numbers of LMA sources. So we will interpret our single probability as the average probability that CRASs could be producing a single LMA source.

With this assumption, we can test the power of our statistical method by performing another Monte Carlo simulation. For each probability to make one additional LMA source that we want to test, we perform many Monte Carlo runs. Each run consists of using the measured B_i for each CRAS and randomly samples a C from our empirical distribution, $f(C|B_i)$. Then,for each CRAS, we sample a uniform random number between 0 and 1. If this number is smaller than the probability that we are testing, then we increase C by 1. Finally, we perform the same statistical analysis described in section 3.1. We perform enough runs that the standard error of the average p value at each average probability is smaller than 10^{-5}.

Table 6. The Average, Standard Deviation, Maximum, and Minimum of the Exclusion Level For 1000 Runs With Randomized CRAS Times

	Average	Standard Deviation	Maximum	Minimum
dt = 2 ms	4.973%	0.193%	5.622%	4.310%
dt = 5 ms	4.470%	0.158%	5.058%	4.039%
dt = 10 ms	4.223%	0.113%	4.668%	3.888%
dt = 25 ms	3.901%	0.104%	4.243%	3.553%

Figure 13 shows the results of our Monte Carlo simulation for dt = 2 ms, for probabilities to produce one additional LMA source between 0 and 0.1. The points in the plot show the average p value, and the bars show the standard deviation of the p values. It is important to note that the bars shown around each point in Figure 13 are not error bars. We notice three important features in this plot. First, when the average probability to produce one LMA source is zero, the p value is uniformly distributed between 0 and 1; particularly, the average p value is 0.5 and the standard deviation is $1/\sqrt{12}$. This shows that our statistical analysis is mathematically consistent. Second, we see that as the average probability to produce a single LMA source increases, our p value decreases to zero, showing that our statistical analysis is, at least on some level, able to distinguish between situations where CRASs do and do not initiate LMA sources. Lastly, we notice that the average p value drops below our alpha value of 0.05 when the average probability to produce a single LMA source exceeds 0.0483. Since the p values of our data are not below 0.05, we can conclude that no more than 4.83% of detected CRASs could have produced LMA sources on time scales of 2 ms. The Monte Carlo results for dt =5 ms, 10 ms, and 25 ms, are very similar, with exclusion levels of 4.31%, 4.26%, and 3.91%.

3.3. Comparison Against a Randomized Data Set

In order to test the consistency of our statistical analysis, we ran another Monte Carlo simulation to test what the results of our analysis would be if the times of the CRASs were assuredly not correlated to times of LMA sources. For each run of this analysis we randomized the time of every detected CRAS individually by picking a random number between 0 and 1 for each CRAS. If the number was 0, then the time of the CRAS was picked to be at the beginning of the 2 s window of LMA data that were processed around that CRAS. If the number was 1, then the time of the CRAS was picked to be at the end of the 2 s window of LMA data, and similarly for numbers in between 0 and 1. Randomized CRAS times picked to be at the beginning and end of the 2 s windows were such that the entire ΔT background bin was still inside of the window of processed LMA data. After randomizing the time of each CRAS individually, we then excluded CRAS that did not have any LMA sources in background or coincident with the new randomized CRAS time. We then performed the same analysis as described above, and we recorded the final value of TS, the p value, the number of CRASs used in the analysis, and the final exclusion level.

We performed 1000 runs for each time scale. Table 3 shows the average, standard deviation, maximum, and minimum of TS for each time scale. Table 4 shows the same statistics for the p values. Table 5 shows the same statistics for the number of CRASs used in the analysis. Finally, Table 6 shows the same statistics for the final exclusion levels. We can see that our results using the measured CRAS times are consistent with the results when the times of the CRAS are randomized.

4. Discussion

We analyzed 4884 recorded CRASs. However, because we are only interested in CRASs during thunderstorms, we only used data from CRAS that have at least one background or coincident LMA source; i.e., there must be at least one LMA source in the ΔT time span. The analyses performed at time scales of dt = 2 ms, 5 ms, 10 ms, and 25 ms, have 325, 501, 651, and 936 CRASs with at least one background or coincident LMA source. The number of analyzed CRAS does not increase linearly with the increase in time because LMA souces tend to be tightly grouped in time. Notably, the ICLRT often begins to collect data hours before a thunderstorm arrives directly overhead, and does not stop collecting data until it is certain that there is no more chance of overhead thunderstorms. As such, there are long periods of time that the ICLRT collects CRAS data without a thunderstorm directly overhead.

Our LMA filtering is strict enough that, if there are LMA sources within 8 km, then there must be dielectric breakdown, hence a thunderstorm, within 8 km. *Pilkey et al.* [2014] showed that the charge regions of thunderstorms around the ICLRT tend to have horizontal extents around 8 km. Thus, since the analyzed CRASs occur during thunderstorm activity and traverse thunderstorm charge regions within 8 km radius (as discussed in section 2.2), then most of the CRASs used in this analysis will have passed directly though a thunderstorm charge region. It is possible that there could be CRASs that have coincident LMA sources but propagated at an angle such that they did not actually pass through the thunderstorm. Since such CRASs could potentially appear to produce LMA sources, but not physically able to do so due to not having passed through a thunderstorm, they will produce some bias toward accepting the hypothesis that CRAS do initiate lightning flashes.

Since we exclude all CRASs that do not have at least one coincident or background LMA source, it is also possible that a CRAS could pass through a thunderstorm but not be included in our analysis. This potentially also creates a bias in our results toward accepting the hypothesis that CRAS do initiate lightning flashes. If a CRAS passes through a thunderstorm when there is no electrical activity and does not initiate a lightning flash, then that CRAS would likely not be included in our analysis. But if a CRAS passes through a thunderstorm and does initiate a flash (within the appropriate time span), then that CRAS will be included in our analysis. The amount of bias will depend upon the ratio of time that there is a strong thunderstorm electric field but no lightning flashes and the amount of time with electrical activity within 8 km. Unfortunately, due to a lack of general understanding of thunderstorm electric fields and a lack of data, it is impossible to guess what that ratio might be.

Another source of bias could be created due to the fact that we only use LMA data within 8 km of the ICLRT. It is possible that a lightning flash could initiate outside of the 8 km, and then propagate into our 8 km radius region. In this situation, we can be fairly confident that any detected CRAS could not have initiated that lightning flash, because our detected CRAS can only initiate lightning flashes within 8 km of the ICLRT (see end of section 2.2). With this in mind, there are three possibilities: (1) the CRAS occurred $\Delta T/2$, or earlier, before the time that the lightning flash propagated into the 8 km radius region, (2) the CRAS occurred within $\pm\Delta T/2$ of the time that the lightning flash propagated into the 8 km diameter region, and (3) the CRAS occurred $\Delta T/2$, or later, after the time that the lightning flash propagated into the 8 km diameter region. In the first possibility, the CRAS will probably not be included in the analysis since the CRAS will likely not have any coincident or background LMA sources. Note that this possibility will not add any bias to our analysis because the CRAS will be excluded independently of whether it did or did not initiate the lightning flash. In the second possibility, there will be a point in time during the ΔT time span when LMA sources suddenly start occurring. This effect will not produce any bias because the time that LMA sources start being detected in 8 km will be random (with respect to the detected CRAS) and so will be accounted for in our background distribution. Finally, the CRAS in the third possibility would also likely be included in our analysis, but will not introduce any bias since typical lightning phenomena, such as lightning leaders, are accounted for in our empirical background distribution function.

While our analysis as a whole does not show any statistical correlation between CRASs and LMA sources, it is possible that there are a few CRASs that are coincident with the initiation of a lightning flash, but do not have enough significance to affect the results of our analysis. Figure 12 shows the LMA sources during the time of the CRAS with the largest contribution to TS. That is, Figure 12 shows the LMA data during the CRAS that deviates the most from our background distribution. This figure shows a large group of LMA sources, probably a cloud-to-ground lightning flash, that starts just before the CRAS and lasts for over 300 ms. However, the fact that there are a number of LMA sources before the CRAS indicates that this CRAS could not have produced this group of LMA sources.

Our analysis shows that we have no evidence that CRASs detected at the ICLRT, which have zenith angles less than 38° and primary energies in between 10^{16} eV and 10^{18} eV, could be producing LMA sources that are strong enough to have chi-square values less than 1, on time scales between 2 ms and 25 ms. Furthermore, our Monte Carlo analysis shows that if the detected CRASs were producing single LMA sources more than 5% of the time, then we should have had statistical significance. For this reason, on average, no more than 5% of CRASs with zenith angles less than 38° and energies between 10^{16} eV and 10^{18} eV can be producing LMA sources. More generally, LMA sources are associated with conventional dielectric breakdown, and so no more than 5% of detected CRASs could have initiated any significant dielectric breakdown.

The time scales of 2 ms and 25 ms are significant because 2 ms is just larger than the average length of the lightning initial breakdown period, and 25 ms is near the average duration of the initial stepped leader in negative cloud-to-ground (CG) lightning (around 35 ms) [*Rakov and Uman*, 2003]. It is possible for our LMA to not capture LMA sources during the short initial breakdown of a lightning flash, and it is conceivable that there are spatially small lightning flashes and phenomena such as narrow bipolar events [*Rison et al.*, 2016] that do not produce a significant number of LMA sources and so cannot be considered by our analysis. However, we observe that most cloud-to-ground and intracloud flashes within 8 km of the ICLRT produce a significant number of LMA sources within a 25 ms time period. Therefore, the probability for a CRAS to produce a single LMA source must be significantly greater than the probability to initiate an entire lightning flash, and we can generalize the results of our analysis and argue that, on average, no more than 5% of detected CRASs could be initiating typical lightning CG or intracloud lightning flashes.

We can extend our results to apply to all CRAS with energies between 10^{16} eV and 10^{18} eV and zenith angles less than 38°, since there should not be any other physical shower parameters, other than zenith angle and primary energy, that both influence the likelihood of a cosmic ray air shower initiating a lightning flash and affects the effective area of our detection network. One possibility of an additional parameter that could influence how our results generalize to the cosmic ray population as a whole is X_{max}. The sensitivity of the ICLRT cosmic ray air shower network will change for different values of X_{max}. This change in sensitivity, however, is already mostly taken into account through primary energy and zenith angle.

Furthermore, since cosmic ray air showers with larger primary energies produce more secondary particles, they may also induce more relativistic runaway electron avalanches. As a result, assuming cosmic ray air showers can initiate lightning, then it seems reasonable that the air showers with larger primary energies might be more efficient at initiating lightning than lower energy air showers. Therefore, since the ICLRT cosmic ray network is sensitive to cosmic rays with primary energies between 10^{16} eV and 10^{18} eV, specifically we found that less than 5% of CRASs with energies lower than 10^{18} eV and zenith angles less than 38° could be initiating lightning flashes.

Lastly, it may be possible to use our results to set an upper limit on the probability that a given lightning flash could have been initiated by a CRAS. This, however, would be a difficult task. It would have to take into account storm area, and lightning flash rates from the storms similar to those used in this work, if not the exact same storms. It would also have to account for the fact that the ICLRT network is only sensitive to a small population of all CRAS, and is not sensitive to CRASs with large zenith angles. As such, we will not attempt to derive an upper limit on the probability that any particular lightning flash was initiated by a CRAS here. We can say, however, that if we integrate equation (2) between energies of 10^{16} eV and 10^{18} eV and zenith angles less than 38°, then we find that the rate of CRASs is around 1.065 CRASs/km^2/min, while the lightning flash rate is around 1.7 min^{-1} for single-peak lightning rate storms that have area around 100 km^2 [*Rakov and Uman*, 2003]. With this simple model, even if every lightning flash were initiated by CRAS, then only about 1.6% of detectable CRASs would initiate lightning; therefore, we cannot exclude the possibility that lightning flashes are initiated by CRAS.

5. Conclusion and Summary

In conclusion, we have compared 4884 cosmic ray air showers, which have primary energies from 10^{16} eV to 10^{18} eV and zenith angles less than 38°, to LMA data recorded at the ICLRT. We have found no evidence that these detected CRASs initiated lightning flashes. Furthermore, we have found that no more than 5% of CRASs with energies less than 10^{18} eV and zenith angles less than 38° could have initiated lightning flashes.

Acknowledgments
This research was supported in part by DOD | Defense Advanced Research Projects Agency (DARPA) grant FA8650-15-604 C-7535 and DOD | Defense Advanced Research Projects Agency (DARPA) contract HR0011-1-10-1-0061. All data in this paper are available from M.A. Uman (uman@ece.ufl.edu).

References

Babich, L. P., E. I. Bochkov, and I. M. Kutsyk (2011), Lightning initiation mechanism based on the development of relativistic runaway electron avalanches triggered by background cosmic radiation: Numerical simulation, *Zh. Eksp. Teor. Fiz.*, *112*(5), 902–909, doi:10.1134/S1063776111040145.

Babich, L. P., E. I. Bochkov, J. R. Dwyer, and I. M. Kutsyk (2012), Numerical simulations of local thundercloud field enhancements caused by runaway avalanches seeded by cosmic rays and their role in lightning initiation, *J. Geophys. Res.*, *117*, A09316, doi:10.1029/2012JA017799.

Chonis, T. G. (2009), Investigating possible links between incoming cosmic rays fluxes and lightning activity over the United States, *Am. Meteorol. Soc.*, *22*, 5748.

Dwyer, J. R. (2005), The initiation of lightning by runaway air breakdown, *Geophys. Res. Lett.*, *32*, L20808, doi:10.1029/2005GL023975.

Dwyer, J. R. (2007), Relativistic breakdown in planetary atmospheres, *Phys. Plasmas*, *14*, 042901, doi:10.1063/1.2709652.

Dwyer, J. R. (2010), Diffusion of relativistic runaway electrons and implications for lightning initiation, *J. Geophys. Res.*, *115*, A00E14, doi:10.1029/2009JA014504.
Dwyer, J. R., and L. P. Babich (2011), Low-energy electron production by relativistic runaway electron avalanches in air, *J. Geophys. Res.*, *116*, A09301, doi:10.1029/2011JA016494.
Fisher, R. A. (1973), *Statistical Methods for Research Workers*, 13th ed., Hafner, New York.
Gaisser, T. K. (1990), *Cosmic Rays and Particle Physics* 1st ed., Cambridge Univ. Press, New York.
Gurevich, A. V., and K. P. Zybin (2005), Runaway breakdown and the mysteries of lightning, *Phys. Today*, *58*(5), 37–43, doi:10.1063/1.199574.
Gurevich, A. V., and A. N. Karashtin (2013), Runaway breakdown and hydrometeors in lightning initiation, *Phys. Rev. Lett.*, *110*, 185005, doi:10.1103/PhysRevLett.110.185005.
Gurevich, A. V., G. M. Milikh, and R. Roussel-Dupré (1992), Runaway electron mechanism of air breakdown and preconditioning during a thunderstorm, *Phys. Lett. A*, *165*, 463–468.
Gurevich, A. V., K. P. Zybin, and R. Roussel-Dupré (1999), Lightning initiation by simultaneous effect of runaway breakdown and cosmic ray showers, *Phys. Lett. A*, *254*, 79–87, doi:10.1016/S0375-9601(99)00091-2.
Gurevich, A. V., Y. V. Medvedev, and K. P. Zybin (2004), New type discharge generated in thunderclouds by joint action of runaway breakdown and extensive atmospheric shower, *Phys. Lett. A*, *329*, 348–361, doi:10.1016/j.physleta.2004.06.099.
Gurevich, A. V., et al. (2009), An intracloud discharge caused by extensive atmospheric shower, *Phys. Lett. A*, *373*(39), 3550–3553, doi:10.1016/j.physleta.2009.07.085.
Hill, D. (2012), The mechanisms of lightning leader propagation and ground attachment, PhD dissertation, Univ. of Florida, Gainesville, Fla. [Available at http://ufdc.ufl.edu/UFE0044602/00001.]
Marshall, T. C., M. P. McCarthy, and W. D. Rust (1995), Electric field magnitude and lightning initiation in thunderstorms, *J. Geophys. Res.*, *100*(D4), 7097–7103.
Milikh, G., and R. Roussel-Dupré (2010), Runaway breakdown and electrical discharges in thunderstorms, *J. Geophys. Res.*, *115*, A00E60, doi:10.1029/2009JA014818.
Pilkey, J. T. (2014), Rocket-triggered lightning propagation and North-Florida thunderstorm charge structure. [Available at http://ufdc.ufl.edu/UFE0047331/00001.]
Pilkey, J. T., M. A. Uman, J. D. Hill, T. Ngin, W. R. Gamerota, D. M. Jordan, J. Caicedo, and B. Hare (2014), Rocket-triggered lightning propagation paths relative to preceding natural lightning activity and inferred cloud charge, *J. Geophys. Res. Atmos.*, *119*, 13,427–13,456, doi:10.1002/2014JD022139.
Rison, W., P. R. Krehbiel, M. G. Stock, H. E. Edens, X. M. Shao, R. J. Thomas, M. A. Stanley, and Y. Zhang (2016), Observations of narrow bipolar events reveal how lightning is initiated in thunderstorms, *Nat. Commun.*, *7*, 10721, doi:10.1038/ncomms10721.
Rakov, V. A., and M. A. Uman (2003), *Lightning: Physics and Effects* 1st ed., Cambridge Univ. Press, Cambridge, U. K.
Solomon, R., V. Schroeder, and M. B. Baker (2001), Lightning initiation-conventional and runaway-breakdown hypotheses, *Q. J. R. Meteorol. Soc.*, *127*, 2683–2704.
Thoudam, S., et al. (2014), LORA: A scintillator array for LOFAR to measure extensive air showers, *Nucl. Instrum. Methods Phys. Res., Sect. A*, *767*, 339–346, doi:10.1016/j.nima.2014.08.021.
Thoudam, S., et al. (2015), Measurement of the cosmic-ray energy spectrum above 10^{16} eV with the LOFAR Radbound Air Shower Array, *Astropart. Phys.*, *73*, 34–43, doi:10.1016/j.astropartphys.2015.06.005.